

## ANALISIS SOAL UAS SEJARAH PEMINATAN KELAS XI IPS SMAN 3 KETAPANG TAHUN AJARAN 2018/2019

**Tri Riski Maulia, Sri Buwono, Ika Rahmatika Chalimi**

Program Studi Pendidikan Sejarah FKIP Untan Pontianak

Email: [tririskimaulia@gmail.com](mailto:tririskimaulia@gmail.com)

### **Abstract**

*This research is to find out the quality of items and the quality of the final exam test in the last semester of historical subject classes XI IPS at SMAN 3 Ketapang School Year 2018/2019. The research method used is a descriptive method with a quantitative approach. The research data was about test items and the student's answer sheet. The analysis used Itean version 3.00 and Microsoft Excel 2010. The research results for the discrimination index of multiple choice are 40 items (89%) is "acceptable" and 5 items (11%) "rejected". As for essay 4 items (80%) is "acceptable" and 1 item (20%) the discrimination index is "rejected". The difficulty index of multiple choice are 3 items (6.7%) is "hard", 23 items (51.1%) is "medium", and 19 items (42.2%) is "easy". As for the essay 5 items (100%) is "hard". The spieeler effectiveness are 12 items (27%) must be revised the deceptions and 33 items (73%) the deceptions is not revised. Item validity of multiple choice are 40 items (89%) valid and 5 items (20%) invalid. Content validity of the multiple choice are 5 items (11%) and 1 essay item (20%) revised or replaced. The reliability index of multiple choice test is 0.771 which means reliable. While the essay is in the number 0.09 which means it is not reliable.*

**Keywords:** *Final Examination, Historical Subject, Item Analysis.*

### **PENDAHULUAN**

Evaluasi dalam pembelajaran berperan penting dalam mendukung terciptanya pembelajaran yang berkualitas. Baik guru maupun peserta didik dapat mengambil pembelajaran dari kegiatan evaluasi yang telah dilakukan. Sunal mengungkapkan bahwa evaluasi adalah sebuah proses yang menggunakan informasi untuk dijadikan sebagai pertimbangan terkait efektivitas suatu program pembelajaran dalam memenuhi kebutuhan siswa. Selain itu, evaluasi juga dapat menjadi acuan untuk tindak lanjut setelah melakukan pengukuran tingkat penguasaan siswa (dalam Susanto, 2016).

Contohnya adalah saat menganalisis butir soal yang akan atau telah diujikan kepada peserta didik. Menurut Nana Sudjana bahwa analisis butir soal berkaitan dengan pengkajian tiap butir soal agar didapatkan gambaran tentang bagaimana kualitas tes yang telah dibuat dan sesuai dengan tujuan

pembelajaran yang telah disusun (dalam Amalia dan Widayati, 2012).

Pada penelitian ini yang dianalisis adalah butir soal ulangan akhir semester sejarah peminatan kelas XI IPS di SMAN 3 Ketapang tahun ajaran 2018/2019. Sebanyak 50 soal yang terbagi atas 45 soal pilihan ganda dan 5 soal essay dianalisis menggunakan program Itean versi 3.00 dan *Microsoft Excel 2010*. Adapun dalam penelitian ini yang dianalisis yaitu kualitas butir soal yang di dalamnya termasuk daya beda, tingkat kesukaran, dan efektivitas pengecoh. Selain itu juga kualitas tes itu sendiri yaitu validitas, dan reliabilitas. Untuk menganalisis butir soal dapat digunakan dua teori, yaitu teori tes klasik dan teori modern, namun pada penelitian ini teori yang digunakan adalah teori tes klasik. Purwanto mengatakan bahwa teori tes klasik sangat terpengaruh pada kedudukan tiap butir tes, atau dengan kata lain sangat

memperhitungkan kedudukan tes. Maka dari itu, butir item memiliki peranan penting karena sedikit kesalahan saat data akan dimasukan, maka besar kemungkinan hasil analisis yang diperoleh akan salah (Purwanto, 2016).

Menurut Allen & Yen bahwa terdapat sedikitnya tujuh asumsi pada teori tes klasik. Asumsi-asumsi inilah yang kemudian memungkinkan dikembangkan formula-formula untuk mengukur atau menganalisis kualitas dari butir yang telah dibuat (dalam Setyawarno, 2017). Dalam teori tes klasik ini ada beberapa tahapan pengukuran yang dilakukan yaitu meliputi daya beda, tingkat kesukaran dan efektivitas pengecoh. Daya beda merupakan kemampuan soal dalam membedakan siswa yang pintar dan siswa yang kurang pintar. Maka dari itu daya beda menjadi acuan utama apakah sebuah butir dapat dikatakan valid atau tidak. Analisis daya beda sendiri bisa menggunakan penafsiran menurut korelasi biserial ataupun korelasi poin biserial, namun pada penelitian ini digunakan korelasi poin biserial. Untuk mencari daya beda soal pilihan ganda dapat digunakan rumus sebagai berikut :

$$r_{pbis} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

Keterangan :

$r_{pbis}$  = koefisien korelasi point biserial

$M_p$  = rata-rata skor subjek yang menjawab benar bagi item yang dicari validitasnya

$M_t$  = rerata skor total

$SD_t$  = standar deviasi dari skor total proporsi

$p$  = proporsi siswa menjawab benar

( $p = \frac{\text{banyaknya siswa yang menjawab benar}}{\text{jumlah keseluruhan siswa}}$ )

$q$  = proporsi siswa menjawab salah

( $q = p-1$ )

(Negara, HRP & Kurniawati, KRA, 2018)

Sedangkan untuk soal essay dapat digunakan rumus yang dikemukakan oleh Arikunto (2013) sebagai berikut :

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\} \{N\sum Y^2 - (\sum Y)^2\}}}$$

Keterangan :

X = skor butir

Y = skor total

$r_{xy}$  = koefisien korelasi antara X dan Y

Selanjutnya adalah untuk menguji daya beda butir dapat dilakukan dengan melihat tingkat kesalahan pengambilan kesimpulan sebesar 5%. Butir yang baik adalah butir yang memiliki tingkat signifikansi ( $p$ ) sebesar 5% ( $p \geq 0,05$ ).

Sebuah butir dikatakan baik apabila butir tersebut memiliki tingkat kesukaran yang sedang, dan setidaknya proporsi soal yang sedang, mudah, dan sukar proporsional. Menurut Crocker dan Algina bahwa tingkat kesukaran dapat didefinisikan sebagai proporsi siswa yang menjawab benar sebuah butir tes (dalam Purwanto, 2016). Berdasarkan pendapat di atas maka TK atau tingkat kesukaran adalah proporsi siswa yang menjawab benar pada tiap butir soal. Selanjutnya untuk menentukan apakah butir tersebut baik atau tidak dapat merujuk pada ketentuan yang telah dibuat. Untuk mencari tingkat kesukaran soal pilihan ganda dapat digunakan rumus yang dikemukakan oleh Purwanto (2016) sebagai berikut :

$$TK = \frac{\sum B}{\sum P}$$

Keterangan :

TK = tingkat kesukaran

SB = jumlah siswa menjawab benar

SP = jumlah siswa peserta tes

Sedangkan untuk soal essay digunakan rumus yang dikemukakan oleh Arifin (2016) sebagai berikut :

$$TK = \frac{X}{SMI}$$

Keterangan :

TK = Tingkat Kesukaran

X = Nilai rata-rata tiap butir soal

SMI = Skor maksimum ideal

Berikut adalah kriteria untuk tingkat kesukaran menurut Arikunto (2013).

**Tabel 1. Kriteria Tingkat Kesukaran**

Rentang Tingkat Kesukaran	Kategori
0,00 – 0,30	Sukar
0,31 – 0,70	Sedang
0,71 – 1,00	Mudah

Selanjutnya untuk mengetahui apakah pengecoh dapat berfungsi dengan baik dapat dilihat pada indeks keefektifan pengecoh yang menurut Arikunto bahwa sebuah pengecoh dapat dikatakan berfungsi dengan baik jika pengecoh tersebut dipilih paling sedikit 5% dari jumlah pengikut tes (Arikunto, 2013).

Tes yang baik juga harus diketahui kualitasnya dengan mengetahui validitas dan reliabilitasnya. Purwanto mengatakan bahwa validitas berkaitan dengan kemampuan sebuah tes dalam mengukur secara tepat untuk apa tes itu digunakan. Ada beberapa tipe dalam validitas yaitu validitas isi (*content*), validitas konstruk (*construct*), dan validitas berdasar kriteria (*criterion-related*) (Purwanto, 2016).

Dalam penelitian ini validitas yang digunakan adalah validitas isi dimana validasi dilakukan oleh ahli. Adapun yang divalidasi adalah kesesuaian instrumen soal yang digunakan dengan kisi-kisi. Arifin mengatakan bahwa validitas kurikuler atau validitas isi ini dapat dilakukan dengan beberapa tahapan, seperti mencocokkan dengan materi tes dengan kisi-kisi atau silabus, berdiskusi antar sesama pendidik, atau mencermati kembali materi tes yang telah dibuat (Arifin, 2016). Namun begitu peneliti juga melakukan pengujian validasi ulang.

Untuk reliabilitas pada penelitian ini dapat dilihat langsung dari pengukuran iteman namun hanya untuk soal pilihan ganda saja. Untuk soal essay reliabilitas dihitung secara manual dengan menggunakan

*Microsoft Excel* 2010 dengan rumus berdasarkan Purwanto (2016) berikut :

$$\alpha_{11} = \left( \frac{n}{n-1} \right) \left( 1 - \frac{\sum S_i^2}{\sum S_t^2} \right)$$

Keterangan :

n = jumlah butir

$\sum S_i^2$  = varians butir

$\sum S_t^2$  = varians total

Menurut Linn dan Kaplan bahwa sebuah tes dapat dikatakan reliabel jika angka yang didapat berada pada angka 0,7 (dalam Widoyoko, 2016).

## METODE PENELITIAN

Dalam penelitian ilmiah diperlukan metode penelitian sebagai cara bagi peneliti untuk mengetahui hal-hal yang akan diteliti. Menurut Syahrudin dan Salim bahwa metode adalah suatu prosedur atau cara untuk mengetahui sesuatu (Syahrudin dan Salim, 2014). Dalam penelitian ini digunakan metode penelitian deskriptif. Adapun data deskripsi yang akan dijabarkan adalah hasil dari analisis butir soal ulangan akhir semester pada penilaian pembelajaran sejarah di kelas XI IPS SMAN 3 Ketapang menggunakan Iteman versi 3.00.

Ratnaningsih mengatakan bahwa Iteman adalah sebuah program atau software yang memungkinkan seseorang untuk menganalisis butir-butir soal dengan menggunakan teori tes klasik. Di dalamnya akan memuat hasil berupa daya beda, tingkat kesukaran, dan statistik sebaran jawaban

berdasarkan pendekatan teori tes klasik (Ratnaningsih, 2011).

Populasi dari penelitian ini adalah seluruh dokumentasi soal ulangan akhir semester mata pelajaran sejarah peminatan kelas XI IPS SMAN 3 Ketapang tahun ajaran 2018/2019 yang berjumlah 45 soal pilihan ganda dan 5 soal essay. Sedangkan sampel penelitiannya adalah seluruh soal ulangan akhir mata pelajaran sejarah peminatan kelas XI IPS SMAN 3 Ketapang tahun ajaran 2018/2019 yang berjumlah 50 soal.

Dalam penelitian ini teknik yang digunakan adalah studi dokumenter. Menurut Dian Pratiwi bahwa teknik studi dokumenter ini adalah sebuah teknik pengumpulan data yang membutuhkan dokumen yang berkenaan dengan data apa yang akan dianalisis (Dian Pratiwi, 2013). Dalam penelitian ini data yang dibutuhkan adalah arsip soal, lembar jawab siswa, kunci jawaban, serta kisi-kisi soal. Selain teknik studi dokumenter, penelitian ini juga melakukan wawancara guna memperoleh informasi lebih dalam kepada guru yang bersangkutan tentang analisis butir soal ini. Dokumen yang diarsipkan berupa lembar soal ulangan akhir semester, skor penilaian, lembar jawaban siswa, kunci jawaban, RPP, serta kisi-kisi soal. Selain itu juga digunakan juga alat berupa pedoman wawancara.

Dalam penelitian ini digunakan teknik analisis data statistik deskriptif, karena peneliti hanya ingin mendeskripsikan data sampel yang telah diuji tanpa membuat kesimpulan. Soal yang dianalisis adalah soal pilihan ganda dan soal essay. Untuk

mempermudah hal tersebut, peneliti menggunakan aplikasi Itaman versi 3.00 untuk menganalisis soal pilihan ganda. Sedangkan untuk soal essay peneliti menggunakan *Microsoft Excel* 2010 untuk menganalisisnya.

## HASIL PENELITIAN DAN PEMBAHASAN

### Hasil

Dari hasil wawancara dengan Ibu Gita Paraheny, S.Pd selaku guru sejarah peminatan diketahui bahwa beliau belum pernah menganalisis butir soal sebelumnya, meski demikian beliau berusaha untuk tetap memperhatikan unsur penting saat membuat soal. Untuk itu saat peneliti melakukan analisis butir soal, beliau ikut serta dengan tujuan untuk mengetahui bagaimana cara menganalisis butir soal yang baik. Seperti yang telah disebutkan di atas bahwa analisis ini terbagi menjadi dua yaitu kualitas butir soal dan kualitas tes. Adapun jenis soal yang dianalisis terdiri dari dua jenis, yaitu pilihan ganda dan essay. Untuk soal pilihan ganda terdapat 45 butir soal, sedangkan untuk soal essay terdapat 5 soal.

Kualitas butir tes dapat dilihat dari daya beda soal pilihan ganda sebanyak 40 soal dari 45 soal daya bedanya “Diterima” dan sebanyak 5 butir soal daya bedanya “Ditolak”. Sedangkan untuk soal essay dapat kita ketahui bahwa 4 dari 5 soal essay daya bedanya “Diterima” dan 1 soal daya bedanya “Ditolak”. Untuk lebih dapat memahaminya dapat dilihat melalui tabel di bawah ini :

**Tabel 2. Hasil Analisis Daya Beda Soal UAS Sejarah Peminatan**

Jenis Tes	Nomor Butir	Jumlah	Keterangan
Pilihan Ganda	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,34,38,39,40,42,43,44,45	40	Diterima
Pilihan Ganda	33,35,36,37,41	5	Ditolak
Essay	1,2,4,5	4	Diterima
Essay	3	1	Ditolak

Sedangkan tingkat kesukaran soal pilihan ganda diketahui bahwa dari 45 butir soal 3 soal tingkat kesukarannya “Sukar”, 23 soal tingkat kesukarannya “Sedang” dan 19 soal tingkat kesukarannya “Mudah”.

Sedangkan untuk soal essay seluruhnya memiliki tingkat kesukaran “Sukar”. Untuk dapat memahaminya dapat dilihat pada tabel berikut :

**Tabel 3. Hasil Analisis Tingkat Kesukaran Soal UAS Sejarah Peminatan**

Jenis Tes	Nomor Butir	Jumlah	Kriteria
Pilihan Ganda	16,21,44	3	Sukar
Pilihan Ganda	1,5,6,9,10,11,13,15,17,18,19,20,24,25,26,28,30,34,35,39,41,42,45	23	Sedang
Pilihan Ganda	2,3,4,7,8,12,14,22,23,27,29,31,32,33,36,37,38,40,43	19	Mudah
Essay	1,2,3,4,5	5	Sukar
Essay	-	0	Sedang
Essay	-	0	Mudah

Terakhir efektivitas pengecoh dapat kita ketahui bahwa sebanyak 12 soal yang pengecohnya harus direvisi dan 33 soal

pengecohnya tidak perlu direvisi. Untuk lebih memahaminya dapat dilihat pada tabel berikut :

**Tabel 4. Hasil Analisis Efektivitas Pengecoh Soal UAS Sejarah Peminatan**

Nomor Butir	Jumlah	Kriteria
1,2,12,14,15,17,23,33,34,37,38,43	12	Revisi
3,4,5,6,7,8,9,10,11,13,16,18,19,20,21,22,24,25,26,27,28,29,30,31,32,35,36,39,40,41,42,44,45	33	Tidak Revisi

Selanjutnya adalah kualitas tes terdiri dari validitas dan reliabilitas. Untuk validitas yang pertama adalah validitas butir soal dimana soal pilihan ganda yang “Valid” ada 40 soal dan 5 soal yang dikatakan “Tidak Valid”. Sedangkan untuk hasil analisis

validitas butir soal essay dapat kita lihat bahwa 4 soal essay dinyatakan “Valid” dan 1 soal “Tidak Valid”. Untuk dapat memahaminya dapat dilihat pada tabel berikut :

**Tabel 5. Hasil Analisis Validitas Butir Soal UAS Sejarah Peminatan**

Jenis Tes	Nomor Butir	Jumlah	Kriteria
Pilihan Ganda	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,34,38,39,40,42,43,44,45	40	Valid
Pilihan Ganda	33,35,36,37,41	5	Tidak Valid

Essay	1,2,4,5	4	Valid
Essay	3	1	Tidak Valid

Selanjutnya untuk mengetahui kualitas instrumen tes yang telah dibuat, maka perlu adanya pengecekan terkait materi, konstruksi, dan bahasa yang dibuat oleh guru. Analisis validitas ini dilakukan oleh dosen Pendidikan Sejarah FKIP UNTAN yaitu Ibu Astrini Eka Putri, M.Pd. Validitas ini dilakukan dengan kisi-kisi sebagai acuan utamanya.

Berdasarkan hasil analisis validasi pada butir soal yang dibuat oleh guru dapat diketahui bahwa untuk aspek materi secara umum mendapat penilaian sangat baik meskipun pada beberapa soal mendapat penilaian cukup baik dan kurang baik pada beberapa indikatornya. Untuk aspek konstruksi secara umum mendapat penilaian sangat baik meskipun pada beberapa soal mendapat penilaian cukup baik, kurang baik, dan tidak baik pada beberapa indikatornya. Sedangkan untuk aspek bahasa secara keseluruhan mendapat penilaian sangat baik.

Meski demikian, pengujian validitas isi juga dilakukan kembali dengan melihat pada hasil analisis yang telah ada, yaitu dengan melihat indeks daya beda butir. Pengujian ini dilakukan sendiri oleh peneliti dengan menandai butir soal yang tidak mampu memenuhi kompetensi yang telah disusun sebelumnya oleh guru. Hasilnya adalah 6 soal tidak valid karena tidak mampu mencapai kompetensi yang ingin dicapai. Hal ini dapat dilihat pada Tabel 5 di atas.

Sedangkan untuk reliabilitas dapat kita ketahui bahwa untuk soal pilihan ganda, reliabilitasnya berada diangka 0,771 yang berarti soal pilihan ganda reliabel. Sedangkan untuk soal essay tingkat reliabilitasnya berada diangka 0,09 yang berarti soal essay tidak reliabel. Untuk dapat memahaminya dapat dilihat pada tabel berikut :

**Tabel 6. Hasil Analisis Reliabilitas Soal UAS Sejarah Peminatan**

Jenis Tes	$\alpha$ (Alpha)	Kriteria
Pilihan Ganda	0,771	Reliabel
Essay	0,089	Tidak Reliabel

### Pembahasan

Berdasarkan hasil penelitian untuk kualitas butir tes yang meliputi dayabeda, tingkat kesukaran, dan efektivitas pengecoh dapat dijabarkan bahwa daya beda memiliki kriteria-kriteria yang menjadi acuan apakah tiap butir item memiliki daya beda yang baik. Adapun kriterianya adalah dengan melihat tingkat kesalahan pengambilan kesimpulan sebesar 5% (Arikunto, 2013). Butir yang baik adalah butir yang memiliki tingkat signifikansi ( $p$ ) sebesar 5% ( $p \geq 0,05$ ).

Untuk daya beda soal pilihan ganda diketahui sebanyak 40 soal (89%) daya bedanya “Diterima” yaitu nomor 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17, 18,19,20,21,22,23,2,25,26,27,28,29,30,31,3 2,34,38,39,40,42,43,44, dan 45. Sedangkan sebanyak 5 soal (11%) daya bedanya “Ditolak” yaitu nomor 33,35,36,37, dan 41. Sedangkan untuk soal essay dapat kita ketahui bahwa 4 soal (80%) daya bedanya “Diterima” yaitu nomor 1,2,4, dan 5. Sedangkan 1 soal (20%) daya bedanya “Ditolak” yaitu nomor 3.

Selanjutnya untuk tingkat kesukaran soal pilihan ganda diketahui soal yang masuk kategori sukar adalah sebanyak 3 butir soal (6,7%) dari keseluruhan soal yang berjumlah 45 soal yaitu nomor 16,21,44. Sedangkan untuk soal yang masuk dalam kategori sedang atau soal yang baik sebanyak 23 soal (51,1%) dari keseluruhan jumlah soal sebanyak 45 butir soal yaitu nomor 1,5,6,9,11,13,15,17,18,19,20,24,25,26,28,30,34,35,39,41,42,45. Sedangkan soal yang termasuk soal mudah sebanyak 19 soal (42,2%) yaitu nomor 2,3,4,7,8,12,14,22,23,27,29,31,32,33,36,37,38,40,43. Untuk soal essay sebanyak 100% atau seluruh soal essay memiliki tingkat kesukaran “Sukar”.

Setelah mencari dan mengetahui nilai dari daya pembeda dan tingkat kesukaran, maka selanjutnya adalah mencari efektivitas pengecoh. Efektivitas pengecoh adalah sejauh mana pengecoh berhasil dipilih baik dari Kelas Atas maupun Kelas Bawah. Efektivitas pengecoh ini hanya berlaku untuk soal dengan tipe pilihan ganda saja (Arikunto, 2013). Berdasarkan hasil penelitian di atas dapat kita lihat bahwa efektivitas pengecoh sebanyak 12 soal (27%) harus direvisi yaitu nomor 1,2,12,14,15,17,23,33,34,37,38, dan 43. Sedangkan sebanyak 33 soal (73%) pengecohnya tidak perlu direvisi yaitu nomor 3,4,5,6,7,8,9,10,11,13,16,18,19,20,21,22,24,25,26,27,28,29,30,31,32,35,36,39,40,41,42,44, dan 45.

Untuk validitas tiap butir soal pilihan ganda soal yang dinyatakan valid sebanyak 40 soal (89%) yaitu nomor 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,25,26,27,28,29,30,31,32,34,38,39,40,42,43,44, dan 45. Sedangkan sebanyak 5 soal (11%) yaitu nomor 33,35,36,37, dan 41 tidak valid. Untuk soal essay 4 soal (80%) yaitu nomor 1,2,4, dan 5 dinyatakan valid, sedangkan soal nomor 3 (20%) dinyatakan tidak valid.

Sedangkan untuk validasi yang dilakukan oleh validator diketahui bahwa

secara umum dari aspek materi, konstruksi, dan bahasa mendapat penilaian sangat baik meski pada beberapa indikator pada setiap aspeknya dinilai cukup baik, kurang baik, dan tidak baik. Namun, hal ini berbeda dengan hasil pengujian validitas isi berdasarkan pada hasil analisis program Iteman versi 3.00. Dari sana dapat dilihat bahwa untuk soal pilihan ganda nomor 33,35,37, dan 41 tidak memiliki daya beda yang baik, sehingga soal-soal tersebut harus direvisi atau diganti dengan catatan kemudian dilakukan uji coba terlebih dahulu agar soal yang dibuat telah memiliki daya beda yang baik. Begitu juga pada soal essay nomor 3, soal tersebut tidak memiliki daya beda yang baik, yang mengakibatkan tidak tercapainya indikator soal yang telah direncanakan sebelumnya.

Terakhir untuk tingkat reliabilitas dapat digunakan pendapat Linn dan Kaplan bahwa batas kritik untuk indeks reliabilitas instrumen adalah 0,7 (dalam Widoyoko, 2016). Berdasarkan analisis soal pilihan ganda dengan menggunakan software Iteman versi 3.00 diketahui bahwa reliabilitas soalnya adalah 0,771 yang berarti soal pilihan ganda ini reliabel. Sedangkan soal essay yang dianalisis menggunakan *Microsoft Excel* 2010 diketahui tingkat reliabilitas soal essay berada di angka 0,09 yang berarti tidak reliabel.

## **SIMPULAN DAN SARAN**

### **Simpulan**

Berdasarkan pada hasil penelitian yang telah dijelaskan di atas maka dapat dilihat bahwa dari kualitas butir soal UAS dari 45 butir soal pilihan ganda ditinjau dari daya beda untuk soal pilihan ganda diketahui bahwa sebanyak 40 soal diterima, dan 5 soal ditolak. Sedangkan untuk soal essay 4 soalnya diterima dan 1 soal ditolak. Adapun tingkat kesukarannya sebanyak 3 soal sukar, 23 soal sedang, dan 19 soal mudah. Sedangkan untuk soal essay seluruh soalnya yang berjumlah 5 soal termasuk sukar. Kemudian efektivitas pengecoh dalam penelitian ini

didapat hasil sebanyak 12 soal harus direvisi dan sebanyak 33 soal tidak perlu direvisi pengecohnya. Untuk validitas butir soal sebanyak 44 soal dinyatakan valid dan 6 soal tidak valid. Untuk validitas tes diketahui bahwa butir soal yang valid baik dari aspek materi, konstruksi, dan bahasa bias dikatakan secara keseluruhan sangat baik meski terdapat beberapa butir soal yang masuk kriteria baik, cukup baik, dan kurang baik pada beberapa indikator di setiap aspeknya. Untuk tingkat reliabilitas diketahui bahwa jenis soal pilihan ganda dinyatakan reliabel dan soal essay tidak reliabel.

### Saran

Menyadari bahwa dalam penelitian ini masih terdapat banyak kekurangan, peneliti bermaksud memberikan beberapa saran yang diharapkan dapat membantu menyempurnakan penelitian-penelitian selanjutnya yang sejenis. Berdasarkan hasil analisis butir soal Ulangan Akhir Semester mata pelajaran Sejarah Peminatan kelas XI IPS di SMAN 3 Ketapang Tahun Ajaran 2018/2019 peneliti menyarankan beberapa hal sebagai perbaikan sebagai berikut : 1.) Bagi guru mata pelajaran sejarah peminatan ada baiknya soal terlebih dahulu diujicobakan kepada siswa, agar guru dapat mengetahui dan memperbaiki soal-soal yang tidak baik, baik dari kualitas butir maupun kualitas dari tes tersebut. 2.) Bagi penelitian selanjutnya diharapkan penelitian saat ini dapat membantu, namun penelitian selanjutnya juga diharapkan untuk memperhatikan kekurangan yang terdapat dalam penelitian ini dan dapat meningkatkan penelitian selanjutnya. 3.) Bagi sekolah diharapkan penelitian ini dapat menjadi pertimbangan untuk lebih memperhatikan kualitas pembelajaran yang telah dilaksanakan di sekolah khususnya pada mata pelajaran sejarah.

### DAFTAR RUJUKAN

Amalia, Ata Nayla & Widayati, Ani. (2012). *Analisis Butir Soal Tes Kendali Mutu Kelas XII SMA Mata*

*Pelajaran Ekonomi Akuntansi Di Kota Yogyakarta Tahun 2012 .X(1) : 4.*

- Arifin, Zainal. (2016). *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- Arikunto, Suharsimi. (2013). *Dasar-Dasar Evaluasi Pendidikan* (Edisi 2). Jakarta: Bumi Aksara.
- Negara, HRP & Kurniawati, KRA. (2018). *Karakteristik Butir Soal Try Out Ujian Nasional SMA Mata Pelajaran Matematika Se-Kota Mataram Dengan Menggunakan Program Iteman*. Jurnal Varian, 1(2): 13.
- Pratiwi, Dian., Zakso, A., Yusuf Ibrahim, M. (2012). *Analisis Kualitas Soal Semester Ganjil Pelajaran Sosiologi Kelas XI SMA*. Jurnal Pendidikan dan Pembelajaran Khatulistiwa 2(9). Retrieved from <http://jurnal.untan.ac.id/index.php/jpdp/article/viewFile/3328/3345>
- Purwanto. (2016). *Evaluasi Hasil Pembelajaran*. Yogyakarta: Pustaka Pelajar.
- Ratnaningsih, Dewi Juliah. (2011). *Analisis Butir Soal Pilihan Ganda Ujian Akhir Semester Mahasiswa di Universitas Terbuka Dengan Pendekatan Teori Klasik*. 12 (2) : 93.
- Setyawarno, Didik. (2017). *Penggunaan Aplikasi Software Iteman (Item and Test Analysis) Untuk Analisis Butir Soal Pilihan Ganda Berdasarkan Teori Tes Klasik*. Jurnal Ilmu Fisika dan Pembelajarannya 1(1). Retrieved from <http://jurnal.radenfatah.ac.id/index.php/jifp/article/download/866/728>
- Susanto, Ahmad. (2016). *Teori Belajar dan Pembelajaran Di Sekolah Dasar Edisi Pertama* (Cetakan ke-4). Jakarta: PRENADAMEDIA GROUP.
- Syahrur & Salim. (2016). *Metodologi Penelitian Kuantitatif*. Bandung: Citapustaka Media.
- Widoyoko, Eko Putra. (2016). *Penilaian Hasil Pembelajaran Di Sekolah* (Cetakan Kedua). Yogyakarta: Pustaka Pelajar.